Cognitive Illusions
Intriguing Phenomena in Thinking, Judgment, and Memory
Rüdiger F. Pohl

Belief bias in deductive reasoning

Publication details
Jonathan St. B. T. Evans, Linden J. Ball, Valerie A. Thompson
Published online on: 04 Mar 2022

How to cite: Jonathan St. B. T. Evans, Linden J. Ball, Valerie A. Thompson. 04 Mar 2022, Belief bias in deductive reasoning from: Cognitive Illusions, Intriguing Phenomena in Thinking, Judgment, and Memory Routledge
Accessed on: 03 Oct 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
10 Belief bias in deductive reasoning

Jonathan St. B. T. Evans, Linden J. Ball, and Valerie A. Thompson

In Chapter 9, work was described on the Wason selection task which has been intensively studied by reasoning researchers and provided evidence of a number of biases. In this chapter, we will look at deductive-reasoning tasks proper, especially those most commonly studied: syllogistic reasoning and conditional inference. These tasks have given rise to widespread claims of belief biases that influence reasoning. One researcher (Stanovich, 1999) even went so far as to claim that the tendency to contextualize all problems is the fundamental computational bias in human cognition, responsible for many errors in human reasoning and decision-making. Before examining these claims, let us look at the nature of deduction itself.

Deductive reasoning involves drawing conclusions that necessarily follow from some given information. For example, if we told you that Sally is shorter than Mary and that Mary is taller than Joan, you could safely conclude that Mary is the tallest of the three. However, if we asked you who was taller, Joan or Sally, you could not infer the answer from the information given. This is because the information given is consistent with two possible situations that you might represent in mental models (Johnson-Laird & Byrne, 1991) as follows:

Mary > Sally > Joan (A)
Mary > Joan > Sally (B)

These models allow us to deduce who is tallest, but not, for example, who is shortest. Most people can solve this kind of problem, although they might need to think about it for a few seconds before answering. Consider a more complex reasoning problem, like the following from the study of Handley and Evans (2000):

You urgently need to get hold of your friend Jane. Jane is on holiday somewhere in Britain. You know she is staying in a youth hostel, but you do not know in which city. Jane, being somewhat mysterious, gives you the following information about her whereabouts:

If Jane is in Hastings, then Sam is in Brighton.
Either Jane is in Hastings or Sam is in Brighton, but not both.

Based on this information, does it follow that:

DOI: 10.4324/9781003154730-12
Belief bias in deductive reasoning

(a) Jane is in Hastings.
(b) Jane is not in Hastings.
(c) It is impossible to tell whether or not Jane is in Hastings?

The reader may care to give this problem some thought before reading on. It is possible to draw a definite conclusion that follows logically from the stated information, although it is hard to see. It involves what is called suppositional reasoning, where you need to suppose a possibility for the sake of argument. In this case, let us suppose that Jane is in Hastings and see what follows. Clearly, we can conclude from the first piece of information that Sam is in Brighton. However, the second statement tells us that either Jane is in Hastings or Sam is in Brighton but not both. So if Jane is in Hastings, by the second statement it follows that Sam is not in Brighton. So we have a contradiction. Our supposition that Jane is in Hastings has led to us to conclude both that Sam is in Brighton and that she is not in Brighton. Since this is an impossible state of affairs, it follows logically that our supposition is false. Hence, we can conclude that Jane is not in Hastings.

This kind of indirect reasoning is very hard for people who are not trained in logic, so don’t worry if you didn’t get the right answer. Handley and Evans gave this problem, among other similar ones, to undergraduate students as a pencil and paper task in a class setting. Only 9.5% offered the correct answer (b). Of the remainder, 48% said (c), impossible to tell, and an astonishing 42.5% gave answer (a) that is the opposite of the correct answer. The authors offered an explanation in terms of mental-model theory. According to this theory, people try to imagine states of affairs, or mental models, that are suggested by the information given. We know that the first statement will suggest the model:

Jane is in Hastings; Sam is in Brighton

even though they may realize that there are other possibilities. When they try to integrate the information in the second statement, they ought to reject this model as it is inconsistent. Those concluding that Jane is in Hastings must have overlooked the significance of the phrase “but not both” in the second statement. However, it is likely that those who did notice the inconsistency mostly moved to the other incorrect answer that it is impossible to tell. Although people may acknowledge that the statement “if p then q” allows possibilities other than p and q to be the case, no other state of affairs comes easily to mind for most people, so it seems to them that no conclusion is possible.

The small number who can solve the above task must either reason by supposition, as shown above, or by examination of all logical possibilities. If you list them out as follows:

(1) Jane is Hastings and Sam is in Brighton.
(2) Jane is in Hastings and Sam is not in Brighton.
(3) Jane is not in Hastings and Sam is in Brighton.
(4) Jane is not in Hastings and Sam is not in Brighton.

and check each in turn against the two statements, you will discover that there is just one state of affairs that is possible, number 3. So you can conclude that Jane is not in Hastings.
and that Sam is in Brighton. However, contrary to proposals once made popular by the
Swiss psychologist Jean Piaget, most adults do not approach deductive reasoning by this
kind of exhaustive logical analysis.

Why is it important to study deductive reasoning in psychology? The answer to this
question has changed quite radically over the past 40 years or so. The paradigm was
developed at a time when Piaget’s views were very influential and when most psychologists
and philosophers saw logic as the basis for rational thinking (Evans, 2002). Hence it
seemed a good idea to give people logical problems and see whether they could solve
them, in order to determine how rationally people can reason. This meant giving people
problems where they must assume the premises are true, introduce none of their real-
world knowledge, and draw only conclusions that strictly and necessarily follow. A large
number of experiments of this kind have been conducted from the 1960s onwards, with
two general findings (Evans, 2007; Manktelow, 2012). First of all, people make many
errors when their answers are compared with a logical analysis of what is right and wrong.
Second, they are highly influenced by the content and context in which the problem is
framed, even though that is irrelevant to the logical task they are set. This led to a big
debate about human rationality (see Chapter 9).

The study of reasoning biases is, however, of considerable psychological interest in its
own right. In this chapter, we will focus mostly on belief bias, a topic which arose origin-
ally in the study of syllogistic reasoning.

**Syllogistic reasoning**

A popular form of reasoning task that is used in the study of deductive reasoning is the
syllogism, originally invented by Aristotle. A syllogism consists of two premises and a con-
clusion, which are always in one of the following four forms:

All A are B  
No A are B  
Some A are B  
Some A are not B

The syllogism always links three terms together. We will use the convention of assuming
that the conclusion links two terms A and C and that each premise therefore connects
with a middle term, B. A complete syllogism might be of the form:

All A are B  
All C are B  
Therefore, all A are C

This argument is a fallacy, that is, its conclusion does not necessarily follow. This can be
seen easily if we substitute some realistic terms for A, B, and C as follows:

All dogs are animals  
All cats are animals  
Therefore, all dogs are cats
The fallacy would be much harder to see if we substituted some different terms as follows:

All tigers are animals
All cats are animals
Therefore, all tigers are cats

People are strongly influenced by whether they agree with conclusions (the belief bias effect, discussed below). However, the task they are set is to say whether or not the conclusion follows in light of the logical structure of the argument. When syllogisms are presented in abstract form — say using letters as in the earlier examples above — error rates are very high indeed. Hundreds of different syllogisms can be formed by varying the terms (technically the mood) used in each premise and conclusion (all, no, some, some not) and by varying the order of reference to the terms A, B, and C. With the conclusion in the form A–C the premises can take four different arrangements (or figures as they are technically known): A–B, B–C; A–B, C–B; B–A, B–C; B–A, C–B. With three statements (two premises and a conclusion) in each of four moods and four different figures, you can make 256 logically distinct syllogisms.

Psychological experiments on abstract syllogistic reasoning are reviewed in detail by Evans et al. (1993, chapter 7) and a study in which people evaluated every possible logical form was reported by Evans et al. (1999). When syllogisms are given to people to evaluate, they frequently say that the conclusion follows, even though the great majority are actually invalid, or fallacious. The endorsement of fallacies is a strong bias in deductive reasoning research as a whole. With syllogisms people are also biased by both the mood and the figure. For example, they are more likely to say that the conclusion follows if it is of a similar mood to the premises. The first syllogism described above, although invalid, has two “all” premises and an “all” conclusion. This type of fallacy is much more often endorsed than one where the conclusion is incongruent with the premises, such as:

All A are B
All C are B
Therefore, no A are C

Note that the conclusion here is possible given the premises, just as it was in the “all” form. People are also biased by the figure of the syllogism. They would, for example, be much more likely to agree with the following argument:

Some A are B
Some B are C
Therefore, some A are C

than this one:

Some A are B
Some B are C
Therefore, some C are A
even though both are actually fallacies. In the first case, the terms seem to follow in a natural order. These fallacies and biases of abstract syllogistic reasoning bear a similarity to the matching bias effect discussed in Chapter 9. They suggest very superficial processing by most of the participants. They also suggest that typical populations of undergraduate students find abstract logical reasoning very difficult. What, however, if problems are made more realistic and easier to relate to everyday knowledge and thinking? How will that affect people’s ability to reason deductively? We consider relevant studies in the following sections.

**Belief bias in syllogistic reasoning**

One of the major phenomena studied in deductive reasoning research is that of belief bias. Belief bias is typically described as a tendency to endorse arguments whose conclusions you believe, regardless of whether they are valid or not. This is not very accurate because, as we have already seen, people tend to endorse fallacies when syllogisms have abstract or neutral content. The belief bias effect is really a suppression of fallacies when conclusions are unbelievable, and so might be better called a debiasing effect, at least for invalid syllogisms.

The usual method by which belief bias is studied involves giving people syllogisms and asking them whether the conclusion necessarily follows (full details of the experimental method can be found in Text box 10.1). Some of the conclusions are logically valid deductions and some are not and some have believable conclusions and some do not. A well-known study by Evans et al. (1983) is often cited as showing clearly the basic phenomena with the necessary controls, although there are much earlier reports of belief bias to be found. Evans et al. presented four categories of syllogisms classified as Valid-Believable (VB), Valid-Unbelievable (VU), Invalid-Believable (IB), and Invalid-Unbelievable (IU), and

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid-Believable</td>
<td>No police dogs are vicious&lt;br&gt;Some highly trained dogs are vicious&lt;br&gt;Therefore, some highly trained dogs are not police dogs</td>
<td>89%</td>
</tr>
<tr>
<td>Valid-Unbelievable</td>
<td>No nutritional things are expensive&lt;br&gt;Some vitamin tablets are expensive&lt;br&gt;Therefore, some vitamin tablets are not nutritional</td>
<td>56%</td>
</tr>
<tr>
<td>Invalid-Believable</td>
<td>No addictive things are inexpensive&lt;br&gt;Some cigarettes are inexpensive&lt;br&gt;Therefore, some addictive things are not cigarettes</td>
<td>71%</td>
</tr>
<tr>
<td>Invalid-Unbelievable</td>
<td>No millionaires are hard workers&lt;br&gt;Some rich people are hard workers&lt;br&gt;Therefore, some millionaires are not rich people</td>
<td>10%</td>
</tr>
</tbody>
</table>
Invalid-Unbelievable (IU). Examples of these are shown in Table 10.1 together with the rates at which people accepted them as valid arguments over three experiments.

The higher acceptance rate for believable syllogisms compared to the unbelievable ones, both for valid and invalid cases, indicates the presence of a strong belief bias. Evans et al. (1983) also showed that people accepted significantly more valid than invalid arguments in line with the logical task set. In addition, they observed a belief by logic interaction such that the belief bias effect was significantly larger for invalid than valid arguments. All three of these effects have been replicated in a number of subsequent studies (see Ball & Thompson, 2018; Evans et al., 1993; Klauer et al., 2000; Newstead et al., 1992).

**Text box 10.1 Classroom demonstration of belief bias in syllogistic reasoning**

**Participants**

The effects are typically quite large and have been demonstrated with small samples. We recommend a minimum of 32 participants drawn from a population of average or above average intelligence.

**Materials**

The material consists of syllogisms like those in Table 10.1. Note that there are two logical forms used. The valid form is

\[
\begin{align*}
\text{No C are B} \\
\text{Some A are B} \\
\text{Therefore, some A are not C}
\end{align*}
\]

and the invalid form is

\[
\begin{align*}
\text{No A are B} \\
\text{Some C are B} \\
\text{Therefore, some A are not C}
\end{align*}
\]

It is necessary to keep these same forms, which were carefully chosen, but the design calls for two syllogisms of each type to be presented. The four shown in Table 10.1 can be used but at least one other set is needed. In order to get a powerful effect, the experimenter needs to make sure that the conclusions follow (believable) or violate (unbelievable) a *class-inclusive* relationship, as do those in the table. For example, all cigarettes are addictive but not all addictive things are cigarettes. So while it is believable, as in the example shown, to say that “some addictive things are not cigarettes” it is unbelievable if it is turned around as “some cigarettes are not addictive”. Ideally, the experimenter also checks the believability of the conclusions by asking a separate group of participants to rate them on a five-point scale from
“Highly Unbelievable” to “Highly Believable” and working out the average ratings. About 16 participants is sufficient for the rating study.

**Design**

The design is within-participants. Each participant will be asked to solve all eight problems, presented in an independently randomized order.

**Procedure**

A booklet with eight pages is given each participant. Each page contains one of the problems, with a layout like this

GIVEN

- No millionaires are hard workers
- Some rich people are hard workers

DOES IT FOLLOW THAT

- Some millionaires are not rich people
- YES/NO

The experimenter assigns each problem a number from 1 to 8 and uses a random-number table or spreadsheet program to work out a separate random sequence for each participant. The problem sheets are numbered in the corner before they are copied and then put together in right order for each participant. The front cover of the booklet can have the written instructions. The participants are instructed to reason deductively, as otherwise the influence of belief could not be considered a bias. Typical instructions (adapted from Evans et al., 1983) would be:

This is a test of reasoning ability. You will be given eight problems. In each case you are given two statements that you should assume to be true. You will also be asked if a further statement – that is, a conclusion – follows from these two statements. If you think the conclusion necessarily follows, then mark the word “YES”, otherwise mark “NO”. Take your time and make sure you have the right answer. Do not go back to a previous problem once you have left it.

**Analysis and results**

There are three effects of interest: an effect of logic, an effect of belief, and an interaction between the two. There are too few measures for parametric analysis, but there is a way around this. First, a table can be made up with a row for each participant and a column for each condition, VB, VU, IB, and IU. For each participant, the number of Yes answers he or she gave for each type is recorded – these must be between 0 and 2. The next step is to compute for each participant the value of
three indices in a further three columns. These indices are computed by adding two different pairs of columns and then subtracting the totals as follows:

\[
\begin{align*}
\text{Logic Index: } & \ (VB + VU) - (IB + IU) \\
\text{Belief Index: } & \ (VB + IB) - (VU + IU) \\
\text{Interaction Index: } & \ (VU + IB) - (VB + IU)
\end{align*}
\]

The first two are fairly obvious. The Logic Index is the number of valid conclusions accepted minus the number of invalid conclusions accepted with belief balanced. Conversely, the Belief Index measures the difference between the acceptance of believable and unbelievable conclusions with logic controlled. The Interaction Index is designed to measure whether (as is usually found) the belief-bias effect is larger for invalid than for valid syllogisms. One might want to think of this as \((IB - IU) - (VB - VU)\) although it is algebraically equivalent to the above. Tests for statistical significance are simple. For each index it can be determined whether it is significantly above zero by using the sign test. For each index the number of participants with a score above zero is counted and compared with the number at or below zero using the binomial test. Since these effects are well known, one-tailed tests can be used in each case.

In the literature, a lot of interest has been focused on the cause of the belief by logic interaction. One explanation offered by Evans et al. (1983) has become known as the Selective Scrutiny Model. According to this model, people tend to accept believable conclusions uncritically and check the logic more thoroughly when conclusions are unbelievable. Hence, they are more likely to detect a fallacy when it leads to a conclusion with which they disagree. A mental-models theory version of this proposes that people form an initial mental model of the premises that supports typically fallacious conclusions which can only be refuted by searching for a counterexample model, that is, one in which the premises do not support the conclusion. Such search for counterexamples is more likely to occur when the conclusion is unbelievable (Oakhill et al., 1989). The Selective Scrutiny Model is supported by evidence that, when people are forced to respond rapidly, the belief bias increases and the belief–logic interaction disappears (Evans & Curtis-Holmes, 2005), presumably because they do not have time to search for counterexamples to the unbelievable conclusion.

**Dual processes and belief bias**

Dual-process theories offer a similar explanation. Recall that this theory posits that people have recourse to two qualitatively different processes: Fast, autonomous (Type 1) processes and slower, working-memory-demanding Type 2 processes (see Chapter 9). On this view, we can suppose that there are two kinds of belief bias: A Type 1 belief bias is a response bias which should favor believable conclusions across the board; a Type 2 belief bias involves motivated reasoning and accounts for the interaction. Since standard dual-process theory claims that Type 1 processing is quicker than Type 2, it follows that the Type 2 belief bias should be suppressed by rapid responding, leading to loss of the interaction.
Some findings in the literature, however, appear inconsistent with the Selective Scrutiny Model. For example, people may spend most time reasoning about the invalid-believable problem, which should be rapidly accepted according to the model (Ball et al., 2006; Thompson et al., 2011). Other authors, however, have interpreted response latencies as consistent with the Selective Scrutiny Model if individual differences in reasoning ability are taken into account when examining participants’ latency profiles (Stupple et al., 2011). A slightly different account is the Selective Processing Model (Evans et al., 2001; Klauer et al., 2000) in which people are supposed only to form one model of the premises, without a subsequent search for counterexamples. In this model, people will try to form a mental model that is consistent with the premises unless the conclusion is unbelievable; in the latter case they will try to find a model which refutes the conclusion, that is, they search for a counterexample model from the start. This is more difficult, however, leading to the interaction usually observed.

Claims have also been made (e.g., Dube et al., 2010) that belief-bias effects in syllogistic reasoning, including the interaction, can be accounted for as a pure response bias when analyzed using the methods of signal detection theory. If this is correct, then belief bias would all be Type 1. However, the argument involves technical issues and has been disputed on theoretical and empirical grounds (Klauer & Kellen, 2011; Trippas et al., 2013). Indeed, Trippas et al. (2014) provide compelling support based on analyses using signal detection theory in favor of a dual-process view of belief-bias effects. In their studies, Trippas et al. presented participants with pairs of syllogisms and asked them to choose which of the two conclusions was valid. The crucial twist in these studies was that for some trials the believability of the two presented conclusions was equated (i.e., they were both unbelievable or were both believable), which controlled for the response-bias component of belief bias. For these trials, however, there was still evidence for the influence of Type 2 motivated reasoning in determining decisions. In related work, Trippas et al. (2015) showed that it is the disposition to think analytically that is the best predictor of motivated reasoning effects in engendering Type 2 belief bias rather than cognitive ability (e.g., IQ or other measures of cognitive capacity).

Although the findings reported by Trippas et al. (2014, 2015) seem to provide convincing support for a dual-process account of belief bias, the issue continues to be contentious (e.g., see Stephens et al., 2019, for a recent defense of a pure response-bias account). Overall, however, our view remains that the balance of evidence supports the involvement of response biases (Type 1 processes) and motivated reasoning (Type 2 processes) in determining belief-bias effects in syllogistic reasoning. This view is also supported by neural-imaging research on belief-bias effects with syllogisms, which has provided good evidence for Type 1 or Type 2 responses being associated with activation in distinct brain regions (e.g., Luo et al., 2013; Tsuji & Watanabe, 2009). Moreover, this view has gained further credence by the proliferation of recent studies of belief bias in syllogistic reasoning that have drawn on dual-process ideas, including studies that have revealed some intriguing findings about the nature of Type 1 and Type 2 processing. We now turn to consider some of these recent findings as they have provoked some important new developments in our understanding of the involvement of dual processes in deductive reasoning.
Logical intuitions

One finding that has sparked considerable interest is the extreme sensitivity that people seem to display to syllogisms in which the logical status of the conclusion conflicts with its belief status. Examining Table 10.1 shows that there are two problems of this type, termed “logic/belief conflict” problems: the Valid-Unbelievable item and the Invalid-Believable item. People’s sensitivity to such problems compared to “no-conflict” problems is revealed in multiple ways, including increased response latencies (e.g., De Neys & Glumicic, 2008; Stupple et al., 2011), heightened autonomic arousal as determined by galvanic skin conductance measures (De Neys et al., 2010), and reduced confidence in responses (De Neys et al., 2011; Thompson et al., 2011). Moreover, this conflict-detection sensitivity is even found in reasoners who respond primarily in accordance with Type 1 beliefs and who do not engage in any motivated Type 2 reasoning (De Neys, 2012, 2014; Stupple et al., 2011).

This latter finding is particularly curious: How can it be that those who are not engaging in slower, Type 2 reasoning still seem to have some awareness of the logical status of the conclusion in terms of its validity or invalidity? In other words, Type 1 response bias should always be completed before an attempt at Type 2 reasoning is initiated, so Type 2 processes should not be able to interfere with Type 1 processes. It is not just the case that the problems that have been presented in such studies have simple and readily discernible logical forms; evidence for people’s sensitivity to the logical status of conclusions with belief/logic conflict problems has also been obtained with moderately complex syllogisms (Trippas et al., 2017) and even with highly complex syllogisms that are responded to with very short response deadlines (Newman et al., 2017), attesting to the robustness of this rapidly occurring conflict-detection effect.

An equally curious finding established by Handley et al. (2011), albeit not with syllogisms but with conditional reasoning problems (discussed below), is that if people are explicitly instructed to respond to conclusions purely in terms of whether they are believable or unbelievable (a reversal of the typical belief-bias paradigm), then the logical status of conclusions is now observed to disrupt effective belief-based responding on logic/belief conflict items (see also Trippas et al., 2017). This finding is important as it suggests that the logical status of a conclusion is somehow determined fast enough to interfere with a normally rapid and straightforward Type 1 belief-based response.

How might the observations that logic/belief conflicts disrupt both logic-based and belief-based responding be explained from a dual-process perspective? A dominant account is that people’s fast, Type 1 processing can be driven by “logical intuitions” (De Neys, 2012, 2014; De Neys & Pennycook, 2019), with the idea here being that people can quickly apprehend the logical status of an argument’s conclusion using implicit, intuitive processes (see Morsanyi & Handley, 2012, for the earliest evidence of people’s apparent intuitive detection of the logic of syllogisms; for further evidence of this effect see Trippas et al., 2016). More recent findings, however, suggest that the concept of logical intuitions may be something of a misnomer. Rather than people possessing and applying a fast intuitive logic, it seems more psychologically plausible to assume that,
when responding either under logic or belief instructions, people are sensitive to fairly superficial structural cues in the presented problems that correlate with the logical status of conclusions (cf. Klauer & Singmann, 2013). The point is that deductive problems have highly formal structures that are likely to trigger a “feeling of rightness” (Ackerman & Thompson, 2017) regarding the validity or invalidity of a given conclusion, thereby influencing responding.

This latter point is not to deny the influence on responding that can arise from the surface features of problems that are correlated with the logical status of conclusion, but is rather to clarify that the basis of the observed effects may reside in available cues in the problem rather than in some kind of deep-seated logical understanding. That said, emerging evidence does suggest that it is people of higher cognitive ability who are best able to pick up on these cues. For example, Thompson et al. (2018) have shown that when reasoners are asked to respond on the basis of belief, then the interference that is caused by the logical status of conclusions is most marked in high-ability reasoners. This suggests that for this group of reasoners a logic-based response is their default response rather than a belief-based one. Lower ability people, on the other hand, show the reverse pattern, where conclusion believability interferes with judgments of validity (Thompson et al., 2018).

A related finding comes from the use of the “two-response paradigm”, in which people are required to give a fast intuitive response and are subsequently given time to deliberate and generate a revised response, if they wish (Thompson et al., 2011). This paradigm has revealed that those who get the answer to deductive reasoning problems correct do so at Time 1, rather than Time 2 (Bago & De Neys, 2017; Thompson et al., 2011). Particularly pertinent to the present discussion, however, is evidence that this tendency is most marked for high IQ people (Raoelison et al., 2020). This finding reinforces the fact that such individuals are more likely to “intuit” the correct answer, again presumably through their ability to pick up on structural cues in presented problems that align with their logical status.

In sum, the emerging picture with respect to syllogistic reasoning is of an intricate interplay between more rapid Type 1 processes and slower Type 2 processes that can bias responses to presented problems in predictable ways under both standard logic instructions and pragmatic instructions to determine the believability of conclusions. Such evidence paves the way toward the development of the next generation of dual-process theories of deduction which will have both the rigor and flexibility to explain a wide range of reasoning phenomena. Having demonstrated that belief bias is a complicated but nevertheless robust effect in the context of syllogistic reasoning, we examine below evidence for belief biases in two other domains: conditional inference, which is the major paradigm used currently in the psychology of deductive reasoning, and also informal inference.

**Conditional inference**

The conditional inference task has been widely employed in the psychology of reasoning for many years, and in recent times has taken over in popularity from the Wason selection task as the most used paradigm. Participants are invited to endorse or reject each of four inferences that can be drawn from a conditional statement. We will illustrate this first with letter-number rules, similar to those used for the abstract version of the Wason selection
task (Chapter 9). The conditional statement relates the letter on one side of the card to a number on the other side. The statement might be

If there is an A one side of the card, then there is a 3 on the other side of the card.

The four inferences are the following:

<table>
<thead>
<tr>
<th>Modus Ponens (MP)</th>
<th>The letter is an A; therefore the number is a 3.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denial of the Antecedent (DA)</td>
<td>The letter is not an A; therefore the number is not a 3.</td>
</tr>
<tr>
<td>Affirmation of the Consequent (AC)</td>
<td>The number is a 3; therefore the letter is an A.</td>
</tr>
<tr>
<td>Modus Tollens (MT)</td>
<td>The number is not a 3; therefore the letter is not an A.</td>
</tr>
</tbody>
</table>

Although there are different theories of what the conditional statement means (Evans & Over, 2004) they all agree that only MP and MT are valid inferences. MP is pretty obvious but MT requires a little reflection. If a card does not have a 3 on it, it cannot have an A because if it did have an A, the number would be 3. With abstract problems like these, most people endorse MP but MT is endorsed significantly less frequently (Evans & Over, 2004, Chapter 3). The DA and AC inferences are fallacies, meaning that they do not necessarily follow: It is consistent with the rule to have a 3 paired with some letter other than A. Studies with abstract conditionals, however, show that these two inferences are quite often endorsed. There is no logical interpretation of the conditional (e.g., as a biconditional), which is consistent with the overall pattern of responding and it is likely that several cognitive biases could operate on particular inferences. As with the matching bias (Chapter 9), some of these can be illustrated by introducing negative components. Consider these two forms of MT:

1. If the letter is A, then the number is 3. The number is not 3, therefore the letter is not A.
2. If the letter is not A, then the number is 3. The number is not 3, therefore the letter is A.

Both inferences are valid, but 2 requires an extra step: Strictly speaking the conclusion of MT is that the letter is not not an A. Logic allows us to remove this double negation as it means the same as the letter is A. Perhaps not surprisingly, fewer people endorse 2 than 1. Now compare these:

3. If the letter is A, then the number is 3. The number is not 3, therefore the letter is not A.
4. If the letter is A, then the number is 3. The number is 4, therefore the letter is not A.

In 4, the negation of the second premise is implicit: The number is a 4 and therefore, by implication, not a 3. In such cases, people draw MT less frequently (Evans & Handley, 1999), an effect analogous to the matching bias discussed in Chapter 9.

The main interest in this chapter is the effect that prior beliefs about realistic conditional statements have on the inferences people draw. Before discussing relevant studies, however, we need briefly to explain the difference between the old and new paradigms in the psychology of reasoning.
Old and new paradigms

The psychology of deductive reasoning has traditionally used the deduction paradigm founded in classical binary logic (Evans, 2002). In classical logic, all statements are either true or false and all inferences either valid or invalid. A valid inference is one whose conclusion can never be false if all of its premises are true. Thus deduction is classically truth-preserving: True premises guarantee true conclusions. The experimental method in the deduction (or old) paradigm reflects this kind of logic. Participants are asked (a) to assume that the premises given are true and (b) to endorse only those conclusions which necessarily follow. Research with this paradigm led to the discovery of many errors and biases in human reasoning. However, a number of researchers began to question whether binary logic and the deduction paradigm provided the right approach for the psychology of reasoning (Evans, 2002; Oaksford & Chater, 1998), leading to the development of a new paradigm which has become popular in recent years.

In making this shift, psychologists were influenced by the writings of some philosophers (e.g., Edgington, 1995) who exposed the problems with conditional statements in the binary-logic paradigm. In order for a conditional always to be true or false, \( \text{if } p \text{ then } q \) has to mean the same thing as \( \neg p \lor q \). This leads to some silly and unacceptable inferences (see Chapter 9). For example, it must be true that “If Boris Johnson is president of the USA, then he will declare war on Russia”, simply because Boris Johnson is not actually president of the USA. In fact, any conditional statement is true in classical logic when the antecedent is false or the consequent is true. So given that 2 + 2 = 4, then we can infer that “If the moon is made of blue cheese, then 2 + 2 = 4” and so on.

The solution favored by Edgington and by many psychologists now (e.g., Evans & Over, 2004) is that a conditional is not always true or false. Instead it has a probability or degree of belief that can only be evaluated when the antecedent is true. To decide if we believe a conditional statement, we perform a thought experiment. We imagine that the antecedent is true and then decide the extent to which we believe the consequent. So if we say to you “If the Democrats are elected, then taxes will rise”, you imagine the premise of this scenario and use your knowledge of the past behavior of the Democrat party, their current policies, the state of the economy, etc. to determine how likely it is that taxes will rise. This determines your belief in the conditional statements, so that \( P(\text{if } p \text{ then } q) = P(q|p) \), often referred to as the Equation. There is now considerable evidence that people’s actual beliefs about conditional statements do conform to the Equation (e.g., Over et al., 2007). People believe conditional statements to the extent that they believe q when they imagine p.

In the new paradigm, participants are no longer asked to assume the truth of premises and are often asked to give degrees of belief in conclusions. With these changes, use of the conditional inference paradigm continues.

Belief bias in conditional inference

Beliefs bias conditional inference but not in the same way as they bias syllogistic reasoning. In syllogistic reasoning, inferences are biased by the believability of the conclusion. With conditional inference, however, it is belief in the conditional statement itself – the major premise – which is important. It was discovered some time ago that the valid inferences, MP and MT, can be suppressed if participants do not believe the conditional statement
Belief bias in deductive reasoning

167

(Byrne, 1991; George, 1995; Stevenson & Over, 1995). Here is an example, using the method employed by Byrne.

**Argument 1**

If Ruth meets her friend, they will go to a play.
Ruth meets her friend.
What follows?

**Argument 2**

If Ruth meets her friend, they will go to a play.
If Ruth has enough money, she will go to a play.
Ruth meets her friend.
What follows?

With Argument 1 most people will say that Ruth goes to the play – Modus Ponens. With Argument 2, however, far fewer people give this conclusion, even though it is as logically necessary as it was in the first argument. The reason is that the second conditional has created doubt about the first one. We cannot be sure that Ruth will go to the play just because she meets her friend; she must have enough money as well. These kinds of pragmatic effects (beliefs induced by context) can be readily demonstrated. It does not need an active suppressor as in the above method. We can simply measure people’s *a priori* belief in conditional statements and observe that fewer inferences are made from the ones they disbelieve. In fact, with this method, all four inferences, valid and invalid, are suppressed by such disbelief (see, e.g., Evans et al., 2010).

The suppression of conditional inferences can be demonstrated using methods of both old and new paradigms. For example, in the classical paradigm people will less frequently say that an inference follows when it is suppressed by disbelief and in the new paradigm they will give a lower probability rating to the conclusion. One study which used both methods (Evans et al., 2010) found that when people were just given arguments and asked to rate probabilities of conclusions (new method) they endorsed fewer inferences of all four kinds regardless of cognitive ability. However, when given standard deductive reasoning instructions (assume premises, decide necessity of conclusion) this belief effect was suppressed by high- but not low-ability participants. This is a result consistent with dual-process theory (see Chapter 9) particularly of the kind proposed by Stanovich (2011). According to Stanovich, people need a combination of cognitive capacity and motivation in order to engage in effective formal reasoning. Unless both are present, people will rely on belief-based reasoning.

The old-paradigm interpretation of these belief effects was that inferences could be blocked by the availability of counterexamples. This permits the binary approach (e.g., mental-model theory). Take the case of AC: “If it is a cat, it has four legs; it has four legs, therefore it is a cat”. We would expect few people to endorse this because a counterexample case, for example, a dog with four legs, is easy to think of. The new paradigm approach would be to say that this conditional statement has a low degree of belief. An important difference, however, is that belief effects are not necessarily regarded as biases in the new paradigm. In this approach, we do not expect people to engage in assumption-based
reasoning except as a special kind of problem solving which only high IQ people seem to be good at. Everyday reasoning, such as that supporting decision-making, can and should take account of the belief we have in premises. The new paradigm still deals with deduction but allows that confidence in conclusions can rightly be affected by confidence in premises (see Evans & Over, 2012).

Belief bias in informal reasoning

In addition to syllogistic and conditional inference studies, psychologists have investigated belief biases in informal reasoning and argumentation. Stanovich and West (1997) devised what they call the Argument Evaluation Task (or AET for short). Here is an example item:

Dale’s belief: 17-year-olds should have the legal right to drink alcoholic beverages.

Dale’s premise or justification for belief: 17-year-olds are just as responsible as 19-year-olds, so they ought to be granted the same drinking rights as other adults.

Critics’ counter-argument: 17-year-olds are three times more likely to be involved in an automobile accident while under the influence of alcohol than 19-year-olds (assume statement factually correct).

Dale’s rebuttal to critics’ counter-argument: 17-year-olds will drink no matter what the law says (assume statement factually correct) so it is useless to try to legislate that they not drink.

Indicate the strength of Dale’s rebuttal to critic’s counter-argument:

A = Very Weak  B = Weak  C = Strong  D = Very Strong

Stanovich and West used a panel of expert judges to decide how strong Dale’s arguments were (in this case, judged to be weak). Of course, participants will also vary in the extent to which they believe the conclusion for which Dale is arguing. If asked to judge the strength of the argument they should disregard such beliefs: Failure to do so is seen as a form of belief bias. Stanovich and West have used the AET in many studies since (Stanovich, 1999) showing that individuals who score well (i.e., resist belief) are both higher in cognitive ability and in the disposition to think rationally, as measured by a number of psychometric scales. Another informal reasoning task measures people’s tendency to take account of the law of large numbers when assessing evidence for propositions they do or do not agree with (Klaczynski & Robinson, 2000). These authors also found evidence of belief bias, for example, disregarding a small sample when the study favored a conclusion with which they agreed. However, the effect was much stronger in older than younger adults.

Although it is tempting to assume that belief biases in formal and informal reasoning tasks must be essentially the same thing, there is little evidence to support this. Thompson and Evans (2012) ran tests of syllogistic belief bias, AET, and the law of large numbers on the same participants. There was little evidence of any correlation between the different measures of belief bias, nor did the tasks respond in the same way to relevant experimental manipulations. As Thompson (2000) has demonstrated previously, many reliable effects in reasoning tasks turn out to be task-specific.
Conclusions

People’s reasoning is undoubtedly and strongly affected by pragmatic factors. When realistic and thematically rich materials are employed, the beliefs that they evoke change the answers that are given to the questions. From the viewpoint of the old paradigm in the psychology of reasoning, all such effects are biases. This is because the old paradigm tests assumption-based reasoning, in which participants must always evaluate the form of a logical argument and assume its premises to be true. Belief effects demonstrate how difficult this is for ordinary people to do. In fact, the only people at all good at this have unusually high IQs.

Indeed, as described above, recent evidence suggests that high-IQ people may form a default response that is based on logic; the logical structure of arguments interferes with their ability to make judgments based on beliefs (Thompson et al., 2018). The accumulating evidence that we discussed about the role of “logical intuitions” in reasoning has challenged modern researchers, and it is not yet clear how to reconcile these findings with the broad base of findings that support the Selective Processing Model. As well, the new paradigm is in its infancy and we have much to learn as yet about whether people generally can take beliefs into account in ways which are effective and appropriate, that is, by weighing degrees of belief against the strength of the evidence or the argument they have been presented. The Thompson and Evans (2012) paper suggested that they do, but there is much to be learned about how well theories derived from formal paradigms can explain more everyday forms of inferences.

Summary

• Belief bias in deductive reasoning is the tendency to draw or approve conclusions that conform with prior beliefs, whether or not logically supported by the premises.
• The classical effect is demonstrated using three-term syllogisms. Many studies have shown three reliable effects: (a) people accept more valid than invalid arguments, (b) people accept more believable than unbelievable conclusions (belief bias), and (c) the belief bias is stronger for invalid arguments.
• These effects are often interpreted within dual-process theory, in which belief bias is attributed to intuitive Type 1 processing, while the validity effect reflects effortful Type 2 reasoning. The belief by logic interaction is interpreted as a Type 2, motivated reasoning bias.
• There have been numerous challenges to this perspective. Although we believe that the broad base of evidence supports this interpretation, it is less clear how to reconcile the dual-process view with the emergent evidence on logical intuitions.
• Belief bias of a different kind can be shown in conditional inference. In this case, it is belief in the conditional statement itself that causes a bias. People are more inclined to accept inferences drawn from believable conditional statements, regardless of their logical validity.
• In the new paradigm of reasoning, influences of belief are not necessarily considered biases. The new approach allows that confidence in conclusions can reflect the degree of belief that people hold in premises. Assumption-based reasoning, required by the old paradigm, is important in formal problem solving but less relevant to everyday decision-making.
• Belief bias has also been demonstrated in informal reasoning tasks, where arguments may be perceived as stronger simply because someone agrees with their conclusions. No clear link with belief bias in formal reasoning tasks has been established, however.

Further reading

Detailed review and discussion of belief-bias models of deductive reasoning is provided by Ball and Thompson (2018). For a less technical and broader treatment of the role of knowledge and belief in reasoning, see Evans (2007, Chapter 4).

References


Belief bias in deductive reasoning


